

09/242054

Practitioner's Docket No. GM50023

CHAPTER II

TRANSMITTAL LETTER
TO THE UNITED STATES ELECTED OFFICE (EO/US)

(ENTRY INTO U.S. NATIONAL PHASE UNDER CHAPTER II)

PCT/US98/05422

20 March 1998

20 March 1997

INTERNATIONAL APPLICATION NO.

INTERNATIONAL FILING DATE

PRIORITY DATE CLAIMED

NOVEL COMPOUNDS

TITLE OF INVENTION

Michael T. Black, et al.

APPLICANT(S)

Box PCT

Assistant Commissioner for Patents

Washington, D.C. 20231

ATTENTION: EO/US

Sir:

I. Applicant herewith submits to the United States Elected Office (EO/US) the following items under 35 U.S.C. § 371:

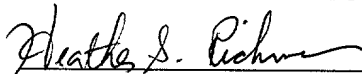
- a. ☒ This express request to immediately begin national examination procedures (35 U.S.C. § 371(f)).
- b. ☒ The U.S. National Fee (35 U.S.C. § 371(c)(1)) and other fees (37 C.F.R. § 1.492) as indicated below:

CERTIFICATION UNDER 37 C.F.R. § 1.10

I hereby certify that this Transmittal Letter and the papers indicated as being transmitted therewith is being deposited with the United States Postal Service on this date **February 8, 1999** in an envelope as "Express Mail Post Office to Addressee" Mailing Label Number **EJ 511 242 063 US**, addressed to the: Assistant Commissioner for Patents, Washington, D.C. 20231.

Heather S. Richman

(type or print name of person mailing paper)



Signature of person mailing paper

2. Fees

Claims Fee	(1) FOR	(2) NUMBER FILED	(3) NUMBER EXTRA	(4) RATE	(5) CALCULATIONS
<input type="checkbox"/> *	TOTAL CLAIMS	26-20=	6	x \$18.00 =	\$ 108.00
	INDEPENDENT CLAIMS	8-3=	5	X 78.00 =	\$ 390.00
	MULTIPLE DEPENDENT CLAIM(S) (if applicable) + \$270.00				
BASIC FEE**	<input checked="" type="checkbox"/> U.S. PTO WAS INTERNATIONAL PRELIMINARY EXAMINATION AUTHORITY Where an International preliminary examination fee as set forth in § 1.482 has been paid on the international application to the U.S. PTO: <input type="checkbox"/> and the international preliminary examination report states that the criteria of novelty, inventive step (non-obviousness) and industrial activity, as defined in PCT Article 33(1) to (4) have been satisfied for all the claims presented in the application entering the national stage (37 CFR 1.492(a)(4))\$96.00 <input checked="" type="checkbox"/> and the above requirements are not met (37 CFR 1.492(a)(1))\$670.00 <input type="checkbox"/> U.S. PTO WAS NOT INTERNATIONAL PRELIMINARY EXAMINATION AUTHORITY Where no international preliminary examination fee as set forth in § 1.482 has been paid to the U.S. PTO, and payment of an international search fee as set forth in § 1.445(a)(2) to the U.S. PTO: <input type="checkbox"/> has been paid (37 CFR 1.492(a)(2))\$790.00 <input type="checkbox"/> has not been paid (37 CFR 1.492(a)(3))\$1,070.00 <input type="checkbox"/> where a search report on the international application has been prepared by the European Patent Office or the Japanese Patent Office (37 CFR 1.492(a)(5))\$930.00				\$ 670.00
	Total of above Calculations				= \$ 1,168.00
SMALL ENTITY	Reduction by 1/2 for filing by small entity, if applicable. Affidavit must be filed also. (note 37 CFR 1.9, 1.27, 1.28)				-
	Subtotal				\$ 1,168.00
	Total National Fee				\$ 1,168.00
	Fee for recording the enclosed assignment document \$40.00 (37 CFR 1.21(h)). (See Item 13 below). See attached "ASSIGNMENT COVER SHEET".				
TOTAL	Total Fees enclosed				\$ 1,168.00

* See Attached Preliminary Amendment Reducing the Number of Claims.

- i. ☐ A check in the amount of _____ to cover the above fees is enclosed.
- ii. ☒ Please charge Account No. 50-0258 in the amount of \$1,168.00. A duplicate copy of this sheet is enclosed.

3. ☐ A copy of the International application as filed (35 U.S.C. § 371(c)(2)):
 - a. ☐ is transmitted herewith.
 - b. ☒ is not required, as the application was filed with the United States Receiving Office.
 - c. ☐ has been transmitted
 - i. ☐ by the International Bureau.
Date of mailing of the application (from form PCT/1B/308):
 - ii. ☐ by applicant on .
Date
4. ☒ A translation of the International application into the English language (35 U.S.C. § 371(c)(2)):
 - a. ☐ is transmitted herewith.
 - b. ☒ is not required as the application was filed in English.
 - c. ☐ was previously transmitted by applicant on .
Date
 - d. ☐ will follow.
5. ☐ Amendments to the claims of the International application under PCT Article 19 (35 U.S.C. § 371(c)(3)):
 - a. ☐ are transmitted herewith.
 - b. ☐ have been transmitted
 - i. ☐ by the International Bureau.
Date of mailing of the amendment (from form PCT/1B/308):
 - ii. ☐ by applicant on (date) .
 - c. ☐ have not been transmitted as
 - i. ☐ applicant chose not to make amendments under PCT Article 19.
Date of mailing of Search Report (from form PCT/ISA/210.):
 - ii. ☐ the time limit for the submission of amendments has not yet expired. The amendments or a statement that amendments have not been made will be transmitted before the expiration of the time limit under PCT Rule 46.1.

6. ☐ A translation of the amendments to the claims under PCT Article 19 (38 U.S.C. § 371(c)(3)):

a. ☐ is transmitted herewith.

b. ☐ is not required as the amendments were made in the English language.

c. ☐ has not been transmitted for reasons indicated at point 5(c) above.

7. ☒ A copy of the international examination report (PCT/IPEA/409)

☐ is transmitted herewith.

☒ is not required as the application was filed with the United States Receiving Office.

8. ☐ Annex(es) to the international preliminary examination report

a. ☐ is/are transmitted herewith.

b. ☐ is/are not required as the application was filed with the United States Receiving Office.

9. ☐ A translation of the annexes to the international preliminary examination report

a. ☐ is transmitted herewith.

b. ☐ is not required as the annexes are in the English language.

10. ☒ An oath or declaration of the inventor (35 U.S.C. § 371(c)(4)) complying with 35 U.S.C. § 115

a. ☐ was previously submitted by applicant on _____ .
Date

b. ☒ is submitted herewith, and such oath or declaration

i. ☐ is attached to the application.

ii. ☒ identifies the application and any amendments under PCT Article 19 that were transmitted as stated in points 3(b) or 3(c) and 5(b); and states that they were reviewed by the inventor as required by 37 C.F.R. § 1.70.

iii. ☐ will follow.

- 4 -

663030-4504250

11. ☒ An International Search Report (PCT/ISA/210) or Declaration under PCT Article 17(2)(a):
- a. ☐ is transmitted herewith.
- b. ☐ has been transmitted by the International Bureau.
Date of mailing (from form PCT/IB/308): .
- c. ☒ is not required, as the application was searched by the United States International Searching Authority.
- d. ☒ will be transmitted promptly upon request.
- e. ☐ has been submitted by applicant on .
Date
12. ☒ An Information Disclosure Statement under 37 C.F.R. §§ 1.97 and 1.98:
- a. ☐ is transmitted herewith.
- Also transmitted herewith is/are:
- ☐ Form PTO-1449 (PTO/SB/08A and 08B).
- ☐ Copies of citations listed.
- b. ☒ will be transmitted within THREE MONTHS of the date of submission of requirements under 35 U.S.C. § 371(c).
- c. ☐ was previously submitted by applicant on .
13. ☐ An assignment document is transmitted herewith for recording.
A separate ☐ "COVER SHEET FOR ASSIGNMENT (DOCUMENT) ACCOMPANYING NEW PATENT APPLICATION" or ☐ FORM PTO 1595 is also attached.
14. ☒ Additional documents:
- a. ☐ Copy of request (PCT/RO/101)
- b. ☒ International Publication No. WO 98/41234.
- i. ☐ Specification, claims and drawing
- ii. ☒ Front page only
- c. ☐ Preliminary amendment (37 C.F.R. § 1.121)

- d. ☒ Other: i. CRF Diskette containing gene sequence listing
ii. Datasheet
iii. Form PCT/IB/301
iv. Form PCT/304

15. ☒ The above checked items are being transmitted
a. ☒ before 30 months from any claimed priority date.
b. ☐ after 30 months.
16. ☐ Certain requirements under 35 U.S.C. § 371 were previously submitted by the applicant on _____, namely:

AUTHORIZATION TO CHARGE ADDITIONAL FEES

- ☒ The Commissioner is hereby authorized to charge the following additional fees that may be required by this paper and during the entire pendency of this application to Account No. 50-0258.
- ☒ 37 C.F.R. § 1.492(a)(1), (2), (3), and (4) (filing fees)
- ☒ 37 C.F.R. § 1.492(b), (c) and (d) (presentation of extra claims)
- ☒ 37 C.F.R. § 1.17 (application processing fees)
- ☐ 37 C.F.R. § 1.17(a)(1)-(5) (extension fees pursuant to § 1.136(a).
- ☐ 37 C.F.R. § 1.18 (issue fee at or before mailing of Notice of Allowance, pursuant to 37 C.F.R. § 1.311(b))
- ☐ 37 C.F.R. § 1.492(e) and (f) (surcharge fees for filing the declaration and/or filing an English translation of an International Application later than 30 months after the priority date).

Respectfully submitted,



ANDREW L. FISH
Attorney for the Applicants
Dechert Price & Rhoads
4000 Bell Atlantic Tower
1717 Arch Street
Philadelphia, PA 19103-2793

Reg. No.: 40,605

Tel. No.: 215/994-2643

Mailing Date: February 8, 1999

FIELD OF THE INVENTION

5

10

15

25

30

Clearly, there is a need for factors, such as the novel compounds of the invention, having a present benefit of being useful to screen compounds for antibiotic activity. Such factors may also be used to determine their role in pathogenesis of infection, dysfunction and disease. There is also a need for identification and characterization of such factors and their

antagonists and agonists which can play a role in preventing, ameliorating or correcting infections, dysfunctions or diseases.

The polypeptides of the invention have amino acid sequence homology to a known amidase from *Synechocystis* sp protein.

5 Each patent application to which this invention claims priority is hereby incorporated by reference in its entirety.

SUMMARY OF THE INVENTION

10 It is an object of the invention to provide polypeptides that have been identified as novel ratA polypeptides by homology between the amino acid sequence set out in Figure 2 [SEQ ID NO:2 or 4] and a known amino acid sequence or sequences of other proteins such as amidase from *Synechocystis* sp protein.

15 It is a further object of the invention to provide polynucleotides that encode ratA polypeptides, particularly polynucleotides that encode the polypeptide herein designated ratA.

In a particularly preferred embodiment of the invention the polynucleotide comprises a region encoding ratA polypeptides comprising the sequence set out in Figure 1 [SEQ ID NO:1 or 3], or a variant thereof.

20 In another particularly preferred embodiment of the invention there is a novel ratA protein from *Streptococcus pneumoniae* comprising the amino acid sequence of Figure 2 [SEQ ID NO:2 or 4], or a variant thereof.

In accordance with another aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Streptococcus pneumoniae* 0100993 strain contained in NCIMB Deposit No. 40794.

25 A further aspect of the invention there are provided isolated nucleic acid molecules encoding ratA, particularly *Streptococcus pneumoniae* ratA, including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

30 In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of ratA and polypeptides encoded thereby.

Another aspect of the invention there are provided novel polypeptides of *Streptococcus pneumoniae* referred to herein as ratA as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

- 5 Among the particularly preferred embodiments of the invention are variants of ratA polypeptide encoded by naturally occurring alleles of the ratA gene.

In a preferred embodiment of the invention there are provided methods for producing the aforementioned ratA polypeptides.

- 10 In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

- 15 In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for (i) assessing ratA expression, (ii) treating disease, for example, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid,, (iii) assaying genetic variation, (iv) and administering a ratA polypeptide or polynucleotide to an organism to raise an immunological response against a bacteria, especially a *Streptococcus pneumoniae* bacteria.

- 20 In accordance with certain preferred embodiments of this and other aspects of the invention there are provided polynucleotides that hybridize to ratA polynucleotide sequences, particularly under stringent conditions.

In certain preferred embodiments of the invention there are provided antibodies against ratA polypeptides.

- 25 In other embodiments of the invention there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in
30 response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided ratA agonists and antagonists, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention there are provided compositions comprising a ratA polynucleotide or a ratA polypeptide for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings depict certain embodiments of the invention. They are illustrative only and do not limit the invention otherwise disclosed herein.

Figure 1 shows the polynucleotide sequence of *Streptococcus pneumoniae* ratA [SEQ ID NO:1 or 3].

Figure 2 shows the amino acid sequence of *Streptococcus pneumoniae* ratA [SEQ ID NO:2 or 4] deduced from the polynucleotide sequence of Figure 1.

GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as the case may be, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988).

Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990)). The well known Smith Waterman algorithm may also be used to determine identity.

Parameters for polypeptide sequence comparison include the following:

1) Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci. USA.* 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program useful with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison WI. The aforementioned parameters are the default parameters for peptide comparisons (along with no penalty for end gaps).

Parameters for polynucleotide comparison include the following:

1) Algorithm: Needleman and Wunsch, *J. Mol Biol.* 48: 443-453 (1970)

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

Gap Length Penalty: 3

Available as: The "gap" program from Genetics Computer Group, Madison WI. These are the default parameters for nucleic acid comparisons.

A preferred meaning for "identity" for polynucleotides and polypeptides, as the case may be, are provided in (1) and (2) below.

(1) Polynucleotide embodiments further include an isolated polynucleotide comprising a polynucleotide sequence having at least a 50, 60, 70, 80, 85, 90, 95, 97 or 100% identity to the reference sequence of SEQ ID NO:1 or 3, wherein said polynucleotide sequence may be identical to the reference sequence of SEQ ID NO:1 or 3 or may include up to a certain integer number of nucleotide alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion, and wherein said

alterations may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of nucleotide alterations is determined by

5 multiplying the total number of nucleotides in SEQ ID NO:1 or 3 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of nucleotides in SEQ ID NO:1 or 3, or:

$$n_n \leq x_n - (x_n \bullet y),$$

10

wherein n_n is the number of nucleotide alterations, x_n is the total number of nucleotides in SEQ ID NO:1 or 3, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is rounded down to

15 the nearest integer prior to subtracting it from x_n . Alterations of a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 or 4 may create nonsense, missense or frameshift mutations in this coding sequence and thereby alter the polypeptide encoded by the polynucleotide following such alterations.

By way of example, a polynucleotide sequence of the present invention may be

20 identical to the reference sequence of SEQ ID NO:2 or 4, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one nucleic acid deletion, substitution, including transition and transversion, or insertion, and wherein said alterations may occur at

25 the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between those terminal positions, interspersed either individually among the nucleic acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of nucleic acid alterations for a given percent identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 or 4 by the integer defining the percent

30 identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2 or 4, or:

$$n_n \leq x_n - (x_n \bullet y),$$

wherein n_n is the number of amino acid alterations, x_n is the total number of amino acids in SEQ ID NO:2 or 4, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_n and y is

5 rounded down to the nearest integer prior to subtracting it from x_n .

(2) Polypeptide embodiments further include an isolated polypeptide comprising a polypeptide having at least a 50,60, 70, 80, 85, 90, 95, 97 or 100% identity to a polypeptide reference sequence of SEQ ID NO:2 or 4, wherein said polypeptide sequence may be identical to the reference sequence of SEQ ID NO:2 or 4 or may include up to a certain integer number of amino acid alterations as compared to the reference sequence, wherein said alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence, and wherein said number of amino acid alterations is determined by multiplying the total number of amino acids in SEQ ID NO:2 or 4 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2 or 4, or:

20

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2 or 4, y is 0.50 for 50%, 0.60 for 60%, 0.70 for 70%, 0.80 for 80%, 0.85 for 85%, 0.90 for 90%, 0.95 for 95%, 0.97 for 97% or 1.00 for 100%, and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a .

By way of example, a polypeptide sequence of the present invention may be identical to the reference sequence of SEQ ID NO:2 or 4, that is it may be 100% identical, or it may include up to a certain integer number of amino acid alterations as compared to the reference sequence such that the percent identity is less than 100% identity. Such alterations are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion, and wherein said alterations may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence or

anywhere between those terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. The number of amino acid alterations for a given % identity is determined by multiplying the total number of amino acids in SEQ ID NO:2 or 4 by the integer defining the percent identity divided by 100 and then subtracting that product from said total number of amino acids in SEQ ID NO:2 or 4, or:

$$n_a \leq x_a - (x_a \bullet y),$$

wherein n_a is the number of amino acid alterations, x_a is the total number of amino acids in SEQ ID NO:2 or 4, y is, for instance 0.70 for 70%, 0.80 for 80%, 0.85 for 85% etc., and \bullet is the symbol for the multiplication operator, and wherein any non-integer product of x_a and y is rounded down to the nearest integer prior to subtracting it from x_a .

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other

reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., *Posttranslational Protein Modifications*:

Perspectives and Prospects, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992) each of which is incorporated by reference herein in its entirety. Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

“Variant(s)” as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

DESCRIPTION OF THE INVENTION

The invention relates to novel ratA polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of a novel ratA gene of *Streptococcus pneumoniae*, which is related by amino acid sequence homology to amidase from *Synechocystis* sp polypeptide. The invention relates especially to ratA having the nucleotide and amino acid sequences set out in Figure 1 [SEQ ID NO:1 or 3]

and Figure 2 [SEQ ID NO:2 or 4] respectively, and to the ratA nucleotide sequences of the DNA deposited in NCIMB Deposit No. 40794 and amino acid sequences encoded thereby.

Deposited materials

A deposit containing a *Streptococcus pneumoniae* 0100993 strain has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (NCIMB), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 April 1996 and assigned NCIMB Deposit No. 40794. The *Streptococcus pneumoniae* strain deposit is referred to herein as "the deposited strain" or as "the DNA of the deposited strain."

The deposited material is a strain that contains the full length ratA DNA, referred to as "NCIMB 40794" upon deposit. The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

The deposit has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

Polypeptides

The polypeptides of the invention include the polypeptide of Figure 2 [SEQ ID NO:2 or 4] (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of ratA, and also those which have at least 70% identity to the polypeptide of Figure 2 [SEQ ID NO:2 or 4] or the relevant portion, preferably at least 80% identity to the polypeptide of Figure 2 [SEQ ID NO:2 or 4], and more preferably at least 90% similarity (more preferably at least 90% identity) to the polypeptide of Figure 2 [SEQ ID NO:2 or 4] and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the polypeptide of Figure 2 [SEQ ID NO:2 or 4] and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with ratA polypeptides fragments may be "free-standing," or comprised within a larger

polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of the amino acid sequence of Figure 2 [SEQ ID NO:2 or 4], or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Streptococcus pneumoniae*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of ratA, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Streptococcus pneumoniae* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

Polynucleotides

Another aspect of the invention relates to isolated polynucleotides that encode the ratA polypeptide having the deduced amino acid sequence of Figure 2 [SEQ ID NO:2 or 4] and polynucleotides closely related thereto and variants thereof.

Using the information provided herein, such as the polynucleotide sequence set out in Figure 1 [SEQ ID NO:1 or 3], a polynucleotide of the invention encoding ratA polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Streptococcus pneumoniae* 0100993 cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as the sequence given in Figure 1 [SEQ ID NO:1 or 3], typically a library of clones of chromosomal DNA of *Streptococcus*

pneumoniae 0100993 in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), which is incorporated by reference herein in its entirety. (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotide set out in Figure 1 [SEQ ID NO:1 or 3] was discovered in a DNA library derived from *Streptococcus pneumoniae* 0100993.

The DNA sequence set out in Figure 1 [SEQ ID NO:1 or 3] contains an open reading frame encoding a protein having about the number of amino acid residues set forth in Figure 2 [SEQ ID NO:2 or 4] with a deduced molecular weight that can be calculated using amino acid residue molecular weight values well known in the art. ratA of the invention is structurally related to other proteins of the rat family, as shown by the results of sequencing the DNA encoding ratA of the deposited strain. The protein exhibits greatest homology to amidase from *Synechocystis* sp protein among known proteins. RatA of Figure 2 [SEQ ID NO:2 or 4] has about 50% identity over its entire length and about 70% similarity over its entire length with the amino acid sequence of amidase from *Synechocystis* sp polypeptide.

The invention provides a polynucleotide sequence identical over its entire length to the coding sequence in Figure 1 [SEQ ID NO:1 or 3]. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker

sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984), incorporated by reference herein in its entirety. Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Streptococcus pneumoniae* ratA having the amino acid sequence set out in Figure 2 [SEQ ID NO:2 or 4]. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having the deduced amino acid sequence of Figure 2 [SEQ ID NO:2 or 4]. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding ratA variants, that have the amino acid sequence of ratA polypeptide of Figure 2 [SEQ ID NO:2 or 4] in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of ratA.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical over their entire length to a polynucleotide encoding ratA polypeptide having the amino acid sequence set out in Figure 2 [SEQ ID NO:2 or 4], and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding ratA polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at

least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Figure 1 [SEQ ID NO:1 or 3].

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein, the disclosure of which is hereby incorporated in its entirety by reference.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in SEQ ID NO:1 or 3 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in SEQ ID NO:1 or 3 or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding ratA and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the ratA gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of the *ratA* gene may be isolated by screening using the DNA sequence provided in SEQ ID NO:1 or 3 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the sequences of SEQ ID NOS:1 and/or 2 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Vectors, host cells, expression

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

- 5 Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard
10 laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic
15 introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and
20 Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from
25 viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector
30 suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL*, (*supra*).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

- 5 Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid
10 chromatography is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

- This invention is also related to the use of the ratA polynucleotides of the invention
15 for use as diagnostic reagents. Detection of ratA in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with an organism comprising the ratA gene may be detected at the nucleic acid level by a variety of techniques.

- 20 Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA or cDNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual,
25 may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled ratA polynucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures.
30 DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers et al., *Science*, 230: 1242 (1985), which is incorporated by reference herein in its entirety. Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage

method. See, e.g., Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1985) which is incorporated by reference herein in its entirety.

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example.

5 For example, RT-PCR can be used to detect mutations. It is particularly preferred to used RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding ratA can be used to identify and analyze mutations. These primers may also be used for amplifying ratA DNA isolated from a sample
10 derived from an individual. The invention further provides these primers with 1, 2, 3 or 4 nucleotides removed from the 5' and/or the 3' end. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious
15 agent.

The invention further provides a process for diagnosing, disease, preferably bacterial infections, more preferably infections by *Streptococcus pneumoniae*, and most preferably otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of
20 cerebrospinal fluid,, comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having the sequence of Figure 1 [SEQ ID NO:1 or 3]. Increased or decreased expression of ratA polynucleotide can be measured using any on of the methods well known in the art for the quantation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting
25 and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of ratA protein compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a ratA protein, in a sample derived from a host are well-known to those of skill in the
30 art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides.

"Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985) which is incorporated by reference herein in its entirety.

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-ratA or from naive libraries (McCafferty, J. *et al.*, (1990), *Nature* 348, 552-554; Marks, J. *et al.*, (1992) *Biotechnology* 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. *et al.*, (1991) *Nature* 352, 624-628).

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against ratA- polypeptide may be employed to treat infections, particularly bacterial infections and especially otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction

between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

- 5 The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH).
- 10 Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

- Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most
- 15 preferably be "humanized"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones, P. et al. (1986), *Nature* **321**, 522-525 or Tempest et al., (1991) *Biotechnology* **9**, 266-273.

- The use of a polynucleotide of the invention in genetic immunization will
- 20 preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff et al., *Hum Mol Genet* 1992, 1:363, Manthorpe et al., *Hum. Gene Ther.* 1993:4, 419), delivery of DNA complexed with specific protein carriers (Wu et al., *J Biol Chem* 1989:264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS*, 1986:83,9551), encapsulation of DNA in various forms of liposomes
- 25 (Kaneda et al., *Science* 1989:243,375), particle bombardment (Tang et al., *Nature* 1992, 356:152, Eisenbraun et al., *DNA Cell Biol* 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., *PNAS* 1984:81,5849).

Antagonists and agonists - assays and molecules

- Polypeptides of the invention may also be used to assess the binding of small
- 30 molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, *e.g.*, Coligan et al., *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of ratA polypeptides or polynucleotides, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to

5 screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising ratA polypeptide and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be a ratA agonist or antagonist. The ability of the candidate molecule to agonize or antagonize the ratA polypeptide is reflected in
10 decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, *i.e.*, without inducing the effects of ratA polypeptide are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter
15 systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in ratA polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for ratA antagonists is a competitive assay that combines ratA and a potential antagonist with ratA-binding molecules, recombinant ratA
20 binding molecules, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. RatA can be labeled, such as by radioactivity or a colorimetric compound, such that the number of ratA molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

25 Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing ratA-induced activities, thereby
30 preventing the action of ratA by excluding ratA from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential

antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of ratA.

5 Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct
10 antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive
15 bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block ratA protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et al.*, *Infect. Immun.* 60:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial ratA proteins that mediate tissue
20 damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example
25 infection of cerebrospinal fluid,.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with ratA, or a fragment or variant thereof, adequate to produce antibody and/ or
30 T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Streptococcus pneumoniae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct

expression of ratA, or a fragment or a variant thereof, for expressing ratA, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a ratA or protein coded therefrom, wherein the composition comprises a recombinant ratA or protein coded therefrom comprising DNA which codes for and expresses an antigen of said ratA or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A ratA polypeptide or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilise the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* Science 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Streptococcus pneumoniae* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the

subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Streptococcus pneumoniae* infection, in mammals, particularly humans.

- 5 The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the
- 10 mouth, mammary glands, urethra or vagina.

- The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or
- 15 intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in
- 20 unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine
- 25 and can be readily determined by routine experimentation.

While the invention has been described with reference to certain ratA protein, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

30 **Compositions, kits and administration**

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for

administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation
5 should suit the mode of administration. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

10 The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion,
15 preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug
20 penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

25 For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of
30 course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves,

pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device.

- 5 Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Streptococcus pneumoniae* wound infections.

- 10 Many orthopaedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

- 15 In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

- 20 Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1µg/ml to 10mg/ml for bathing of wounds or indwelling devices.

- A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5µg/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

EXAMPLES

- 30 The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1 Strain selection, Library Production and Sequencing

The polynucleotide having the DNA sequence given in SEQ ID NO:1 or 3 was obtained from a library of clones of chromosomal DNA of *Streptococcus pneumoniae* in *E.*

coli. The sequencing data from two or more clones containing overlapping *Streptococcus pneumoniae* DNAs was used to construct the contiguous DNA sequence in SEQ ID NO:1 or 3. Libraries may be prepared by routine methods, for example:

Methods 1 and 2 below.

- 5 Total cellular DNA is isolated from *Streptococcus pneumoniae* 0100993 according to standard procedures and size-fractionated by either of two methods.

Method 1

- 10 Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 11kbp in size are rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library packaged by standard procedures and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

Method 2

- 15 Total cellular DNA is partially hydrolyzed with a one or a combination of restriction enzymes appropriate to generate a series of fragments for cloning into library vectors (e.g., RsaI, PstI, AluI, BshI235I), and such fragments are size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E.coli* infected with the packaged library. The library is amplified by standard procedures.
- 20

What is claimed is:

1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:
 - (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising amino acids of SEQ ID NO:2 or 4;
 - (b) a polynucleotide which is complementary to the polynucleotide of (a); and
 - (c) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a) or (b).
2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.
3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.
4. The polynucleotide of Claim 2 comprising the nucleotide sequence set forth in SEQ ID NO:1 or 3.
5. The polynucleotide of Claim 2 comprising nucleotide sequence set forth in SEQ ID NO:1 or 3.
6. The polynucleotide of Claim 2 which encodes a polypeptide comprising amino acid sequence of SEQ ID NO:2 or 4.
7. An isolated polynucleotide comprising a member selected from the group consisting of:
 - (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding the same mature polypeptide expressed by the ratA gene contained in NCIMB Deposit No. 40794;
 - (b) a polynucleotide complementary to the polynucleotide of (a); and
 - (c) a polynucleotide comprising at least 15 bases of the polynucleotide of (a) or (b).
8. A vector comprising the DNA of Claim 2.
9. A host cell comprising the vector of Claim 8.
10. A process for producing a polypeptide comprising: expressing from the host cell of Claim 9 a polypeptide encoded by said DNA.
11. A process for producing a cell which expresses a polypeptide comprising transforming or transfecting the cell with the vector of Claim 8 such that the cell expresses the polypeptide encoded by the cDNA contained in the vector.
12. A process for producing a ratA polypeptide or fragment comprising culturing a host of claim 9 under conditions sufficient for the production of said polypeptide or fragment.

13. A polypeptide comprising an amino acid sequence which is at least 70% identical to amino acid sequence of SEQ ID NO:2 or 4.
14. A polypeptide comprising an amino acid sequence as set forth in SEQ ID NO:2 or 4.
15. An antibody against the polypeptide of claim 13.
16. An antagonist which inhibits the activity of the polypeptide of claim 13.
17. A method for the treatment of an individual having need of ratA comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 13.
18. The method of Claim 16 wherein said therapeutically effective amount of the polypeptide is administered by providing to the individual DNA encoding said polypeptide and expressing said polypeptide *in vivo*.
19. A method for the treatment of an individual having need to inhibit ratA polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 16.
20. A process for diagnosing a disease related to expression of the polypeptide of claim 13 comprising:
determining a nucleic acid sequence encoding said polypeptide.
21. A diagnostic process comprising:
analyzing for the presence of the polypeptide of claim 13 in a sample derived from a host.
22. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 13 comprising:
contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;
and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.
23. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with ratA, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said animal from disease.

24. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of ratA fragment or a variant thereof, for expressing ratA, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.

25. An immunological composition comprising a DNA which codes for and expresses a ratA polynucleotide or protein coded therefrom which, when introduced into a mammal, induces an immunological response in the mammal to a given ratA polynucleotide or protein coded therefrom.

26. A polynucleotide consisting essentially of a DNA sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in SEQ ID NO:1 or 3 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence set forth in SEQ ID NO:1 or 3 or a fragment thereof; and isolating said DNA sequence.

FIGURE 1(a). [SEQ ID NO:1]

ATGACTTTTAAACAATAAAACTATTGAAGAGTTGCACAACTCCTTGTCTCTAAGGAAATT
TCTGCAACAGAAATTGACCAAGCAACACTTGAAAAATATCAAGTCTCGTGAGGAAGCCATC
AATTCAATTGTACCATCGCTGAGGAGCAAGCTTTGTTCAAGCTAAAGCCATTGATGAA
GCTGGAATTGATGCTGACAAATGTCCTTCAGGAATTCCAATTGCTTTAAGGATAACATC
TCTACAGACGGTATTCTCACAACTGCTGCCCTCAAAAATGCTCTACAACTATGAGCCAATC
TTTGTATCGACAGCTGTTGCCAATGCGGTGTTCAAGGAAACCAAAAGGCATGATTGTCGTTGGAAAGACC
AACATGGACGAATTTGCTATGGGTGGTTCAGGTGAAACTTCACACTACGAGCAACTAAA
AACGCTTGGGACCACAGCAAGGTTCTGTGGTCACTCAAGTGGTTCAGTGGTTCGCGCAGCTGTA
GCCTCAGGACAAAGTTCGCTTGTCACTTGGTCTGATCTGATCTGATGTTCCATCCGCCAACCT
GCTGCCCTTCAACGGAACTCGTTGGTCTCAAAACCACTACGGAACAGTTTCACGTTTCGGT
CTCAATTGCCCTTTGGTAGCTCATTAGACAGATTGGACCTTTTGCTCCTACTACTTCTGCTCCT
AATGCCCTCTTGCTCAACGCTATTGCCAGCGAAGATGCTAAAGACTCTACTTCTGCTCCT
GTCCGCATCGCCGACTTTACTTCAAAAATCGGCCAAGACATCAAGGGTATGAAAAATCGCT
TTGCCTTAAGGAATACCTCGGTGAAGGAATTAAACCCAGAGGTTAAGGAAACCATTTCTAAAT
GCCGCTAAACACTTTGAAAATTTGGGTGCTATTGTCGAAGAGTCAGCCTTCCCTCACTCT
AAATACGGAGTTGCCGTATACATCATCGCTTCATCAGAACTTCATCAAACTTGCAA
CGCTTCGACGGTATCCGTTACGGCTATCGCGCAGAAAGTCAACCACTTGATGAAATC
TATGTAAACAGCCGAAGCCAAAGTTTGGTGAAGAGTGAAGCGCCGTATCATGCTGGGT
ACTTTCAGTCTTTCATCAGGTTACTACGATGCCCTACTATAAGAAGGCTGGACAGGTCCGT
ACACTTATCATCAAGATTTGAAAAAGTCTTCGCGGATTACGATTGATTTTGGGTCCA
ACTGCTCCAAAGTTGCCATAGACTTGGATTCTCTCAACCATGACCCAGTTGCCATGTAC
TTAGCCGACCTATTGACCAATACCTGTAACTTGGCAGGACTGCTGGAATTTTCGATTCCCT
GCTGGATTCTCTCAAGGCTACCTGTGCGACTCCAATTGATTGGTCCCAAGTACTCTGAG
GAAACCATTTACCAAGCTGCTGCTTTTGAAGCAACAACAGACTACCAACAACAACAA
CCCCGTGATTTTGGAGGTGACAAAC

6680213" 45024260

FIGURE 1(b). [SEQ ID NO:3]

TTTGCAACAGAAATTGACCCCAAGCAACACACTTGAAATAATCAAGTCTCGTGAGGAAGCCATCAATTCAATTG

FIGURE 2(a). [SEQ ID NO:2]

MTFNKTIIEELHNLVSKHSATELTQATLENIKSREEAINSFVTIAEEQALVQAKAIDE
AGIDADNVLSGIPLAVKDNISTDGIILTTAASKMLYNYEPIFDATAVANAKTKGMI VVGKT
NMDEFAMGGSGETSHYGATKNAWDHSKVPGGSSSGSAAA VASGQVRLSLGSDTGGsirQP
AAFNGIVGLKPTYGTVSRFGLIAFGSSLDQIGPFAPTIVENALLLNAIASEDAKDSTSAP
5 VRIADFTSKIQQDIKGMKIALPKEYLGEINPEVKETILNAAKHFEKLGAIVEEVSLPHS
KYGVAVYYIIASSEASSNLQRFDGIRYGYRAEDATNLDEIYVNSRSQGFGEVKKRRIMLG
TFSLSSGYDAYKKAGQVRTLI IQDFEKVFADYDLILGPTAPSVAYDLDLSLNHDPVAMY
LADLLTIPVNLAGLPGISIPAGFSQGLPVGLQLIGPKYSEETIYQAAAAFEATTDYHKQQ
PVIFGGDN

668020"45024260

FIGURE 2(b). [SEQ ID NO:4]

FATELTQATLENKSRREEANSFVTIAEEQALVQAKAIDEAGIDADNVLSGIPLAVKDNISTDGILTTAA

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

" Novel Compounds "

the specification of which (check one)

☐ is attached hereto.

☒ was filed on **20 March 1998** as Serial No. **PCT/US98/05422**
and was amended on (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below any foreign application for patent or Inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Number	Country	Filing Date	Priority Claimed
--------	---------	-------------	------------------

I hereby claim the benefit under Title 35, United States Code, Section 119(e) of any United States provisional application(s) listed below.

Application Number	Filing Date
60/041,131	20 March 1997
60/041,130	20 March 1997

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application.

Serial No.	Filing Date	Status
------------	-------------	--------

I hereby appoint the following attorneys and/or agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Stephen Venetianer	Registration No. <u>25,659</u>
Janice E. Williams	Registration No. <u>27,142</u>
James M. Kanagy	Registration No. <u>29,550</u>
Edward T. Lentz	Registration No. <u>30,191</u>
William T. King	Registration No. <u>30,954</u>
Linda E. Hall	Registration No. <u>31,763</u>
Mary E. McCarthy	Registration No. <u>32,917</u>
Charles M. Kinzig	Registration No. <u>33,252</u>
Dara L. Dinner	Registration No. <u>33,680</u>
Patricia A. Schreck	Registration No. <u>33,777</u>
Yuriy P. Stercho	Registration No. <u>33,797</u>
Kirk Baumeister	Registration No. <u>33,833</u>
Wayne J. Dustman	Registration No. <u>33,870</u>
Jeffrey A. Sutton	Registration No. <u>34,028</u>
William T. Han	Registration No. <u>34,344</u>
Nora Stein-Fernandez	Registration No. <u>36,689</u>
Alissa M. Eagle	Registration No. <u>37,126</u>
Soma G. Simon	Registration No. <u>37,444</u>
Edward R. Gimmi	Registration No. <u>38,891</u>
Zoltan Kerekes	Registration No. <u>38,938</u>
Andrew L. Fish	Registration No. <u>40,605</u>
Arthur E. Jackson	Registration No. <u>34,354</u>
Allen Bloom	Registration No. <u>29,135</u>
Kara W. Swanson	Registration No. <u>39,851</u>

Address all correspondence and telephone calls to Authur Jackson, Dechert Price & Rhoads at 4000 Bell Atlantic Tower, 1717 Arch Street, Philadelphia, Pennsylvania 19103-2793, whose telephone number is 215-994-4000.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full Name of Inventor: Michael Terence BLACK

Inventor's Signature: 

Date: 22/6/98

Residence: 502 Milhouse Way, Chester Springs, Pennsylvania 19425 ^{PA}

Citizenship: United Kingdom

Post Office Address: SmithKline Beecham Corporation
Corporate Intellectual Property - UW2220
P.O. Box 1539
King of Prussia, Pennsylvania 19406-0939

200 Full Name of Inventor: Elizabeth Jane LAWLOR

Inventor's Signature: E. Lawlor

Date: 22 June 1998

Residence: 32 Cours de Vincennes, 75020 Paris France ^{FRX}

Citizenship: United Kingdom

Post Office Address: SmithKline Beecham Corporation
Corporate Intellectual Property - UW2220
P.O. Box 1539
King of Prussia, Pennsylvania 19406-0939

300 Full Name of Inventor: Ceri John LEWIS

Inventor's Signature: Cemi

Date: 30th June 1998

Residence: 4 Shepherds Hall, Linton Cambridgeshire, United Kingdom

Citizenship: United Kingdom ^{GBX}

Post Office Address: SmithKline Beecham Corporation
Corporate Intellectual Property - UW2220
P.O. Box 1539
King of Prussia, Pennsylvania 19406-0939

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: SmithKline Beecham Corporation
- (ii) TITLE OF INVENTION: Novel Compounds
- (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: Dechert, Price & Rhoads
 - (B) STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
 - (C) CITY: Philadelphia
 - (D) STATE: PA
 - (E) COUNTRY: USA
 - (F) ZIP: 19103-2793
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: Windows 95
 - (D) SOFTWARE: FastSEQ for Windows Version 2.0b
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER: 60/041,130
 - (B) FILING DATE: 20-MAR-1997
-
- (A) APPLICATION NUMBER: 60/041,131
 - (B) FILING DATE: 20-MAR-1997
- (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Falk, Stephen T
 - (B) REGISTRATION NUMBER: 36,795
 - (C) REFERENCE/DOCKET NUMBER: GM50023

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: 215-994-2488
- (B) TELEFAX: 215-994-2222
- (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1464 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGACTTTTA	ACAATAAAAC	TATTGAAGAG	TTGCACAATC	TCCTTGCTCTC	TAAGGAAATT	60
TCTGCAACAG	AATTGACCCA	AGCAACACTT	GAAAATATCA	AGTCTCGTGA	GGAAGCCATC	120
AATTCATTG	TCACCATCGC	TGAGGAGCAA	GCTCTGTTC	AAGCTAAAGC	CATTGATGAA	180
GCTGGAATTG	ATGCTGACAA	TGTCCTTTCA	GGAATTCCAC	TTGCTGTTAA	GGATAACATC	240
TCTACAGACG	GTATTCTCAC	AACTGCTGCC	TCAAAAATGC	TCTACAACCTA	TGAGCCAATC	300
TTTGATGCGA	CAGCTGTTGC	CAATGCAAAA	ACCAAAGGCA	TGATTGTCGT	TGGAAAGACC	360
AACATGGACG	AATTTGCTAT	GGGTGGTTCA	GGTGAACTT	CACACTACGG	AGCAACTAAA	420
AACGCTTGGG	ACCACAGCAA	GGTTCCTGGT	GGGTCATCAA	GTGGTTCTGC	CGCAGCTGTA	480
GCCTCAGGAC	AAGTTCGCTT	GTCACTTGGT	TCTGATACTG	GTGGTTCCAT	CCGCCAACCT	540
GCTGCCTTCA	ACGGAATCGT	TGGTCTCAAA	CCAACCTACG	GAACAGTTTC	ACGTTTCGGT	600
CTCATTGCCT	TTGGTAGCTC	ATTAGACCAG	ATTGGACCTT	TTGCTCCTAC	TGTTAAGGAA	660
AATGCCCTCT	TGCTCAACGC	TATTGCCAGC	GAAGATGCTA	AAGACTCTAC	TTCTGCTCCT	720
GTCCGCATCG	CCGACTTTAC	TTCAAAAATC	GGCCAAGACA	TCAAGGGTAT	GAAAATCGCT	780
TTGCCTAAGG	AATACCTCGG	TGAAGGAATT	AACCCAGAGG	TTAAGGAAAC	CATTCTAAAT	840
GCCGCTAAAC	ACTTTGAAAA	ATTGGGTGCT	ATTGTGGAAG	AAGTCAGCCT	TCCTCACTCT	900
AAATACGGAG	TTGCCGTATA	CTACATCATC	GCTTCATCAG	AAGCTTCATC	AAACTTGCAA	960
CGCTTCGACG	GTATCCGTTA	CGGCTATCGC	GCAGAAGATG	CAACCAACCT	TGATGAAATC	1020
TATGTAAACA	GCCGAAGCCA	AGGTTTTGGT	GAAGAAGTGA	AGCGCCGTAT	CATGCTGGGT	1080
ACTTTCAGTC	TTTCATCAGG	TTACTACGAT	GCCTACTATA	AGAAGGCTGG	ACAGGTCCGT	1140
ACACTTATCA	TTCAAGATTT	CGAAAAAGTC	TTGCGGGATT	ACGATTGAT	TTTGGGTCCA	1200
ACTGCTCCAA	GTGTTGCCTA	TGACTTGGAT	TCTCTCAACC	ATGACCCAGT	TGCCATGTAC	1260
TTAGCCGACC	TATTGACCAT	ACCTGTAAAC	TTGGCAGGAC	TGCCTGGAAT	TTGATTCTCT	1320
GCTGGATTCT	CTCAAGGTCT	ACCTGTCGGA	CTCCAATTGA	TTGGTCCCAA	GTACTCTGAG	1380
GAAACCATTT	ACCAAGCTGC	TGCTGCTTTT	GAAGCAACAA	CAGACTACCA	CAAACAACAA	1440
CCCGTGATTT	TTGGAGGTGA	CAAC				1464

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 488 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

Met	Thr	Phe	Asn	Asn	Lys	Thr	Ile	Glu	Glu	Leu	His	Asn	Leu	Leu	Val
1				5					10					15	
Ser	Lys	Glu	Ile	Ser	Ala	Thr	Glu	Leu	Thr	Gln	Ala	Thr	Leu	Glu	Asn
			20					25					30		
Ile	Lys	Ser	Arg	Glu	Glu	Ala	Ile	Asn	Ser	Phe	Val	Thr	Ile	Ala	Glu
			35					40					45		
Glu	Gln	Ala	Leu	Val	Gln	Ala	Lys	Ala	Ile	Asp	Glu	Ala	Gly	Ile	Asp
			50				55					60			
Ala	Asp	Asn	Val	Leu	Ser	Gly	Ile	Pro	Leu	Ala	Val	Lys	Asp	Asn	Ile
65						70					75				80
Ser	Thr	Asp	Gly	Ile	Leu	Thr	Thr	Ala	Ala	Ser	Lys	Met	Leu	Tyr	Asn
				85					90					95	
Tyr	Glu	Pro	Ile	Phe	Asp	Ala	Thr	Ala	Val	Ala	Asn	Ala	Lys	Thr	Lys
				100					105				110		
Gly	Met	Ile	Val	Val	Gly	Lys	Thr	Asn	Met	Asp	Glu	Phe	Ala	Met	Gly
				115				120					125		
Gly	Ser	Gly	Glu	Thr	Ser	His	Tyr	Gly	Ala	Thr	Lys	Asn	Ala	Trp	Asp
				130				135				140			
His	Ser	Lys	Val	Pro	Gly	Gly	Ser	Ser	Ser	Gly	Ser	Ala	Ala	Ala	Val
145						150					155				160
Ala	Ser	Gly	Gln	Val	Arg	Leu	Ser	Leu	Gly	Ser	Asp	Thr	Gly	Gly	Ser
				165					170					175	
Ile	Arg	Gln	Pro	Ala	Ala	Phe	Asn	Gly	Ile	Val	Gly	Leu	Lys	Pro	Thr
				180					185				190		
Tyr	Gly	Thr	Val	Ser	Arg	Phe	Gly	Leu	Ile	Ala	Phe	Gly	Ser	Ser	Leu
				195				200				205			
Asp	Gln	Ile	Gly	Pro	Phe	Ala	Pro	Thr	Val	Lys	Glu	Asn	Ala	Leu	Leu
				210			215					220			
Leu	Asn	Ala	Ile	Ala	Ser	Glu	Asp	Ala	Lys	Asp	Ser	Thr	Ser	Ala	Pro
225						230					235				240
Val	Arg	Ile	Ala	Asp	Phe	Thr	Ser	Lys	Ile	Gly	Gln	Asp	Ile	Lys	Gly
				245						250				255	
Met	Lys	Ile	Ala	Leu	Pro	Lys	Glu	Tyr	Leu	Gly	Glu	Gly	Ile	Asn	Pro

260 265 270
 Glu Val Lys Glu Thr Ile Leu Asn Ala Ala Lys His Phe Glu Lys Leu
 275 280 285
 Gly Ala Ile Val Glu Glu Val Ser Leu Pro His Ser Lys Tyr Gly Val
 290 295 300
 Ala Val Tyr Tyr Ile Ile Ala Ser Ser Glu Ala Ser Ser Asn Leu Gln
 305 310 315 320
 Arg Phe Asp Gly Ile Arg Tyr Gly Tyr Arg Ala Glu Asp Ala Thr Asn
 325 330 335
 Leu Asp Glu Ile Tyr Val Asn Ser Arg Ser Gln Gly Phe Gly Glu Glu
 340 345 350
 Val Lys Arg Arg Ile Met Leu Gly Thr Phe Ser Leu Ser Ser Gly Tyr
 355 360 365
 Tyr Asp Ala Tyr Tyr Lys Lys Ala Gly Gln Val Arg Thr Leu Ile Ile
 370 375 380
 Gln Asp Phe Glu Lys Val Phe Ala Asp Tyr Asp Leu Ile Leu Gly Pro
 385 390 395 400
 Thr Ala Pro Ser Val Ala Tyr Asp Leu Asp Ser Leu Asn His Asp Pro
 405 410 415
 Val Ala Met Tyr Leu Ala Asp Leu Leu Thr Ile Pro Val Asn Leu Ala
 420 425 430
 Gly Leu Pro Gly Ile Ser Ile Pro Ala Gly Phe Ser Gln Gly Leu Pro
 435 440 445
 Val Gly Leu Gln Leu Ile Gly Pro Lys Tyr Ser Glu Glu Thr Ile Tyr
 450 455 460
 Gln Ala Ala Ala Ala Phe Glu Ala Thr Thr Asp Tyr His Lys Gln Gln
 465 470 475 480
 Pro Val Ile Phe Gly Gly Asp Asn
 485

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTTGCAACAG AATTGACCCA AGCAACACTT GAAAATATCA AGTCTCGTGA GGAAGCCATC 60
 AATTCATTG 70

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 70 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Phe Ala Thr Glu Leu Thr Gln Ala Thr Leu Glu Asn Ile Lys Ser Arg
 1              5              10              15
Glu Glu Ala Ile Asn Ser Phe Val Thr Ile Ala Glu Glu Gln Ala Leu
      20              25              30
Val Gln Ala Lys Ala Ile Asp Glu Ala Gly Ile Asp Ala Asp Asn Val
      35              40              45
Leu Ser Gly Ile Pro Leu Ala Val Lys Asp Asn Ile Ser Thr Asp Gly
      50              55              60
Ile Leu Thr Thr Ala Ala
65              70

```